

Original Paper

# Machine Learning Approach to Reduce Alert Fatigue Using a Disease Medication–Related Clinical Decision Support System: Model Development and Validation

Tahmina Nasrin Poly<sup>1,2,3</sup>, MSc; Md.Mohaimenul Islam<sup>1,2,3</sup>, MSc; Muhammad Solihuddin Muhtar<sup>2</sup>, BSc; Hsuan-Chia Yang<sup>1,2,3</sup>, PhD; Phung Anh (Alex) Nguyen<sup>2,4</sup>, PhD; Yu-Chuan (Jack) Li<sup>1,2,3,5,6</sup>, MD, PhD

<sup>1</sup>Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

<sup>2</sup>International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

<sup>3</sup>Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

<sup>4</sup>Department of Healthcare Information & Management, Ming Chuan University, Taoyuan City, Taiwan

<sup>5</sup>Department of Dermatology, Wan Fang Hospital, Taipei, Taiwan

<sup>6</sup>TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei, Taiwan

**Corresponding Author:**

Yu-Chuan (Jack) Li, MD, PhD

Graduate Institute of Biomedical Informatics

College of Medical Science and Technology

Taipei Medical University

15 F, No. 172-1, Sec. 2, Kellung Rd, Da'an dist

Taipei, 106

Taiwan

Phone: 886 0966546813

Email: [jaak88@gmail.com](mailto:jaak88@gmail.com)

## Abstract

**Background:** Computerized physician order entry (CPOE) systems are incorporated into clinical decision support systems (CDSSs) to reduce medication errors and improve patient safety. Automatic alerts generated from CDSSs can directly assist physicians in making useful clinical decisions and can help shape prescribing behavior. Multiple studies reported that approximately 90%-96% of alerts are overridden by physicians, which raises questions about the effectiveness of CDSSs. There is intense interest in developing sophisticated methods to combat alert fatigue, but there is no consensus on the optimal approaches so far.

**Objective:** Our objective was to develop machine learning prediction models to predict physicians' responses in order to reduce alert fatigue from disease medication–related CDSSs.

**Methods:** We collected data from a disease medication–related CDSS from a university teaching hospital in Taiwan. We considered prescriptions that triggered alerts in the CDSS between August 2018 and May 2019. Machine learning models, such as artificial neural network (ANN), random forest (RF), naïve Bayes (NB), gradient boosting (GB), and support vector machine (SVM), were used to develop prediction models. The data were randomly split into training (80%) and testing (20%) datasets.

**Results:** A total of 6453 prescriptions were used in our model. The ANN machine learning prediction model demonstrated excellent discrimination (area under the receiver operating characteristic curve [AUROC] 0.94; accuracy 0.85), whereas the RF, NB, GB, and SVM models had AUROCs of 0.93, 0.91, 0.91, and 0.80, respectively. The sensitivity and specificity of the ANN model were 0.87 and 0.83, respectively.

**Conclusions:** In this study, ANN showed substantially better performance in predicting individual physician responses to an alert from a disease medication–related CDSS, as compared to the other models. To our knowledge, this is the first study to use machine learning models to predict physician responses to alerts; furthermore, it can help to develop sophisticated CDSSs in real-world clinical settings.

(*JMIR Med Inform* 2020;8(11):e19489) doi: [10.2196/19489](https://doi.org/10.2196/19489)

**KEYWORDS**

clinical decision support system; alert fatigue; machine learning; artificial neural network

## Introduction

Initiation of computerized provider order entry (CPOE) systems has allowed physicians to order medications, laboratory tests, and other ancillary services electronically [1]. CPOE systems create an opportunity to improve patient care by decreasing medication errors, reducing redundant test orders, and promoting standardized clinical practice [2,3]. However, CPOE is often integrated with a clinical decision support system (CDSS) in order to make better clinical decisions through guidance, alerts, and reminders. A CDSS is always combined with software algorithms that generate alerts during orders entered into a CPOE by physicians [4,5]. Each of these alerts addresses a meaningful clinical issue relevant to the administration process and has a positive impact on identifying and preventing erroneous or less optimal prescription [6-8].

The productivity of CDSSs is often impaired by generating distracting alerts in the system (ie, a high volume of clinically irrelevant alerts) [9,10]. van der Sijs et al [11] suggested that an ideal CDSS should have high specificity and sensitivity, provide clear information, and facilitate safe and efficient handling of alerts. A recent study reported that approximately 90%-95% of medication alerts are overridden by providers [12,13], and more than half of overrides are due to alerts being deemed clinically irrelevant [14]. The main concern is that these large numbers of clinically irrelevant alerts might cause alert fatigue and consume too much time and mental energy. Moreover, it sometimes leads staff to override both critical warnings and unimportant alerts. Getting frequent false alerts can desensitize physicians so that providers always ignore and mistrust alerts with acceleration [15]. Ignoring clinically relevant alerts too much triggers patient harm and is associated with an increased rate of mortality.

Until now, significant efforts and strategies have been implemented in minimizing alert fatigue, such as the administration of highly specific algorithms [16], customization of third-party providers' sets of alerts [17], and execution of

tiered severity grading to stratify and lessen the number of false alerts [18]. Several studies suggested turning off frequently overridden alerts [19], updating clinical content to deliver the most current evidence at the point of care, and holding consensus meetings between physicians and pharmacists [20]. Since physicians increasingly adopt electronic prescribing, the progression and proclamation of CDSS alerts might depend, in part, on whether providers find medication safety alerts valuable.






Machine learning is comprised of a collection of techniques that has the potential to learn complex rules and to identify patterns from multidimensional datasets. It has been effectively employed in many areas, such as disease risk prediction [21], classification [22], and health care utilization [23]. To our knowledge, no studies have examined machine learning techniques regarding medication alert reduction in a large number of alert analyses among physicians of different specialties. We hypothesized that machine learning models could predict physician responses, which would ultimately directly assist in developing a sophisticated CDSS for reducing alert fatigue. Therefore, the primary objective of this study was to develop and validate machine learning models to reduce alert fatigue by predicting physician responses. This study may provide perspective on the perceived usefulness of CDSS alerts in patient care and insights into how to design better alert systems in real-world clinical settings. It can contribute to minimizing the number of alerts in the user interface, ensuring the appropriate prescription, and reducing the severity of unintended consequences.

## Methods

### Ethical Approval and Study Process

This type of study does not require Institutional Review Board review, following the policy of the National Health Research Institutes in Taiwan, as it provides a large amount of computerized, deidentified data. The entire study process is shown in Figure 1.

**Figure 1.** Study design process. ATC: Anatomical Therapeutic Chemical classification system; AUROC: area under the receiver operating characteristic curve; CDSS: clinical decision support system; EHR: electronic health record; ICD: International Classification of Diseases.

 Data Collection	 Data Preprocessing	 Features Selection	 Model Development	 Model Evaluation
<p>Source: EHR-integrated CDSS Duration: 10 months</p>	<ul style="list-style-type: none"> <li>❖ Duplication removal</li> <li>❖ Mapping</li> <li>❖ Used 3 digits for ICD and 5 digits for ATC</li> </ul>	<ul style="list-style-type: none"> <li>❖ Consultation with experts</li> <li>❖ Automated by machine learning</li> </ul>	<ul style="list-style-type: none"> <li>❖ Training (80%)</li> <li>❖ Testing (20%)</li> <li>❖ Validation (20% from training set)</li> </ul>	<ul style="list-style-type: none"> <li>❖ Accuracy</li> <li>❖ AUROC</li> <li>❖ Sensitivity and Specificity</li> <li>❖ Positive and negative predictive value</li> </ul>

### Data Source

We collected data from an electronic health record (EHR)-integrated disease medication-related CDSS from a university teaching hospital in Taiwan. We considered only prescriptions that generated alerts due to a prescription error in

the CDSS. The data collection period was between August 2018 and May 2019. During the 10-month study period, 9213 prescriptions generated alerts that accounted for approximately 3% of total prescriptions provided by physicians.

## Data Preprocessing

The first step of this study was to clean the data. In the dataset, lots of duplications of prescriptions appeared, which means there were several prescriptions with the same patient's registration number, diagnosis code (ie, International Classification of Diseases, 10th Revision, Clinical Modification [ICD-10-CM]), and drug code (ie, Anatomical Therapeutic Chemical [ATC] classification system code). Therefore, we removed those prescriptions and kept the most recent prescription. A total of 6453 prescriptions were considered to develop machine learning-based prediction models. A prescription with the Taiwan National Health Insurance code as the diagnosis code was mapped to the ICD-10-CM code. Data normalization was carried out by converting all the values between 0 and 1. Finally, the data were converted into a matrix that included the diagnosis code, drug code, department ID, and physician ID.

## Feature Selection

There could be more than 20 different clinical variables available in a single prescription. Therefore, feature selection is essential in order to keep the variables within a manageable size to be able to optimize the prediction model. The feature selection process was completed in three stages: (1) consultation with an expert (YL) who is a physician and specialist in CDSSs, (2) automated feature selection via machine learning algorithms, and (3) reduction of the number of input variables by using only the first three digits of the diagnosis code (ie, ICD-10-CM) and the first five digits of the drug code (ie, ATC). The patient's age, the patient's gender, the diagnosis code (ie, ICD-10-CM), the drug code (ie, ATC), the physician ID, and the department ID were considered as input variables. We then created a matrix for the diagnosis code (ie, ICD-10-CM), the drug code (ie, ATC code), the physician ID, and the department ID. A total of 6453 input variables were used to develop a machine learning model with binary outcomes.

**Table 1.** List of input variables.

Variable	Input column contents	Input column number
Patient's gender	Male or female	1
Diagnosis code	First 3 digits of the ICD-10-CM <sup>a</sup> code	822
Drug code	First 5 digits of the ATC <sup>b</sup> classification system code	262
Physician ID	Physician ID	227
Department ID	Department ID	29

<sup>a</sup>ICD-10-CM International Classification of Diseases, 10th Revision, Clinical Modification.

<sup>b</sup>ATC: Anatomical Therapeutic Chemical.

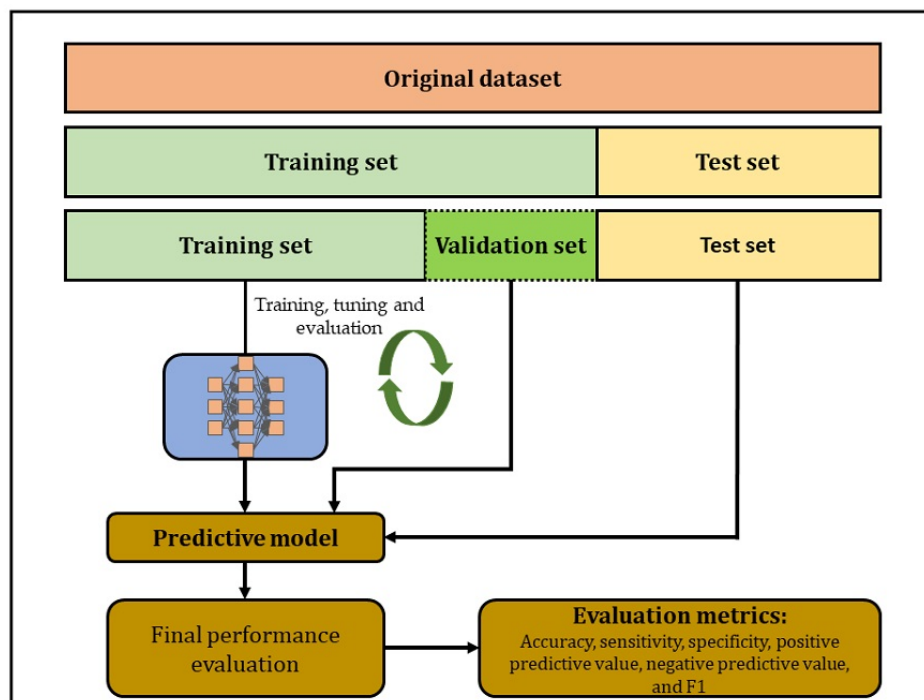
## Model Development

### Overview

The objective of the model was to reliably predict what would be physicians' responses to an alert. We divided the entire dataset into two parts: (1) the training dataset (80% of the dataset) and (2) the testing dataset (20% of the dataset).

However, the model was trained using 60% of the dataset as the training set and 20% of the dataset as an internal validation set. The remaining 20% of the dataset was used for testing our model's performance (see [Figure 2](#)). Model development was carried out using Python 3.6 software (Python Software Foundation). Python is a free and open-source programming language and environment for statistical computing and graphics.

**Figure 2.** Distribution of training and testing datasets for model development.



### Artificial Neural Network

Artificial neural networks (ANNs) were first introduced in the 1940s; recently, they have become extremely powerful and one of the most popular machine learning models that interconnects with adaptive simple processing elements. They usually work by mimicking the biological nervous systems responsible for knowledge processing and knowledge representation [24]. ANN-based algorithms have already shown high performance in terms of accuracy, sensitivity, and specificity for classification problems. Therefore, the application of ANNs has increased globally in recent years in health care research, including in drug development, pattern recognition, disease prediction, disease diagnosis, and disease prognosis. ANNs consist of three layers of neurons: the *input layer*, the *hidden layers*, and the *output layer*. The hidden layer can be a single or multiple layer. Every hidden layer is comprised of an activation function. In our study, we used three hidden layers, with the *rectified linear unit* (ReLU) activation function in the first and second hidden layers, and the *sigmoid* activation function in the third hidden layer.

The ReLU is a widely used activation function in the prediction model. It converts input values from 0 to  $\alpha$ . In the third layer, we used a sigmoid activation function due to a nonlinear nature. The sigmoid function is also one of the most commonly used activation functions for binary classification. The sigmoid

function converts output classes between 0 and 1. The ANN was designed to be a classification model that can predict the responses from multiple physicians while minimizing prediction errors by using *binary cross-entropy* as loss function and the stochastic gradient descent method for optimization. Moreover, 100 epochs were used in the ANN model where maximum accuracy and minimum loss for training and validation can be achieved.

### Random Forest

Random forest (RF) is also known as *ensemble learning* because it is an ensemble of a large number of individual decision trees [25]. Each tree in the RF model spits out a class prediction, and the class with the most votes becomes our model's prediction. However, RF applies to both the *classification* and *regression* models.

### Naïve Bayes

Naïve Bayes (NB) is a classification model that uses the *Bayesian probability* theory during prediction [26]. It is also known as a probabilistic classifier. In 1960, the NB model was first introduced for text classification by the text retriever community [27]. However, there are several types of NB algorithms for parameter estimation and event models, such as *Gaussian naïve Bayes*, *multinomial naïve Bayes*, and *Bernoulli naïve Bayes*. Bayes theorem is expressed as equation 1 in Figure 3.

**Figure 3.** Equations. FN: false negative; FP: false positive; NPV: negative predictive value; PPV: positive predictive value; TN: true negative; TP: true positive.

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (1)$$

$$X = (x_1 + x_2 + x_3, \dots, x_n) \quad (2)$$

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)} \quad (3)$$

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y) \quad (4)$$

$$\hat{A} = \operatorname{argmin} E_{x,y} [L(y, A(x))] \quad (5)$$

$$\hat{A}(x) = \sum_{i=1}^M y_i h_i(x) + \text{constant} \quad (6)$$

$$h(x_i) = \begin{cases} +1 & \text{if } w \cdot x + b \geq 0 \\ -1 & \text{if } w \cdot x + b < 0 \end{cases} \quad (7)$$

$$\left[ \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(w \cdot x_i - b)) \right] + \lambda \|w\|^2 \quad (8)$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (9)$$

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \quad (10)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \quad (11)$$

$$PPV = \frac{(\text{Sensitivity} \times \pi)}{(\text{Sensitivity} \times \pi + (1 - \text{Specificity}) \times (1 - \pi))} \quad (12)$$

$$NPV = \frac{(\text{Specificity} \times (1 - \pi))}{(\text{Specificity} \times (1 - \pi) + (1 - \text{Sensitivity}) \times \pi)} \quad (13)$$

The variable  $y$  is the class variable that represents whether the alert will be accepted or rejected given the condition. Variable  $X$  represents the features like drugs, disease, and demographic.  $X$  is given as equation 2 in Figure 3.

Here,  $x_1, x_2 \dots x_n$  represent the features (ie, they can be mapped to outcome: accept or reject alert). By substituting for  $X$  and expanding using the chain, the rule is given in equation 3 in Figure 3. In our model, the class variable  $y$  has two outcomes: accept or reject. There could be cases where the classification is multivariable. Therefore, the equation 4 in Figure 3 is used to find the class variable  $y$  with maximum probability.

### Gradient Boosting

Gradient boosting (GB) is one of the promising machine learning algorithms that has already shown better prediction for classification [28]. It can be used both in classification and regression models. Like RF, GB is a set of decision trees, but the main differences are how the trees are built and how the results are combined. In the RF model, each tree is built independently, while in the GB model they are built one tree at a time. The GB model works in a forward stage-wise manner and converts weak learners to strong learners [29]. The most interesting part of the GB algorithm is that it can easily fit into the new model. Moreover, the RF model combines results at the end of the process, by averaging or *majority rules*, while the GB model combines results along the way [30].

In the training set, input variables such as drugs and diseases, make a set  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  of known values of  $x$  and corresponding values of  $y$ . The goal is always to find an approximation  $\hat{A}(x)$  to find a function  $A(x)$  that minimizes the expected value of the specified loss function  $L(y, A(x))$ , as shown in equation 5 in Figure 3.

The GB model assumes a real-valued  $y$  and calculates an approximation  $\hat{A}(x)$  in the form of a weighted sum of functions  $h_i(x)$  from  $H$  classes, which are called base or weak learners, as shown in equation 6 in Figure 3.

### Support Vector Machine

Support vector machine (SVM) is a supervised machine learning algorithm. SVM is used both in *classification* and *regression* problems [31]. It is also used to solve linear and nonlinear problems and works well for many complex problems. The idea of SVM is simple: it creates a line or a hyperplane that separates the data into classes. The hypothesis function  $h$  is defined as shown in equation 7 in Figure 3.

The point above or on the hyperplane is classified as a class +1, and the point below the hyperplane is classified as a class -1. The SVM classifier works in the form shown in equation 8 in Figure 3.

## Model Performances

### Overview

To evaluate the performance of five machine learning algorithms, we calculated accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic curve (AUROC). For calculating those measures, we measured true positive, true negative, false positive, and false negative. The definitions of the six parameters are given below.

### Accuracy

Accuracy is the test by which we can see how accurate our model is. The equation to calculate accuracy is shown in equation 9 in [Figure 3](#).

### Sensitivity

Sensitivity is the test by which we can determine a positively identified case. The equation to calculate sensitivity is shown in equation 10 in [Figure 3](#).

### Specificity

Specificity is the measure by which we can measure correctly identified cases from negative cases. The mathematical equation is given in equation 10 in [Figure 3](#).

### PPV and NPV

PPV and NPV are two basic measures in biomedical studies. PPV is the probability that the positively identified case is positive. The mathematical equation for PPV is given in equation 12 in [Figure 3](#). Similarly, the NPV is the probability that the negatively identified case is negative. The mathematical equation is given in equation 13 in [Figure 3](#).

### AUROC

AUROC is a performance measure by which we can evaluate the performance of the model. AUROC is a performance matrix

for *discrimination*; it shows the predictive model's ability to discriminate between positive and negative cases.

## Results

### Dataset Characteristics

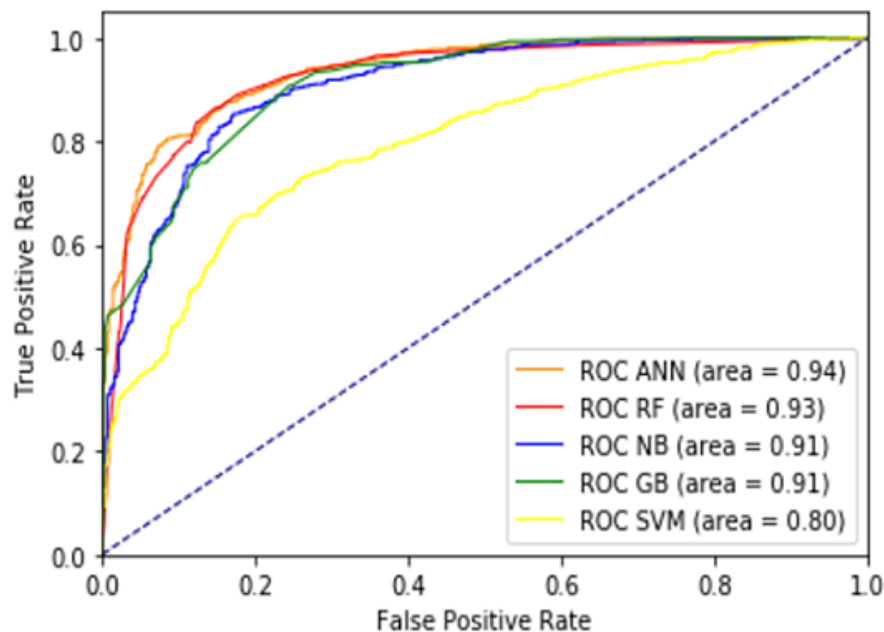
A total of 9214 prescriptions with an alert were collected during the 10-month study period. After preprocessing and removing duplicate prescriptions with the same registration numbers, 6453 prescriptions were used to develop our models. The neurology department got the highest number of alerts (1039/6453, 16.10%). Of those alerts, 546 (52.55%) were accepted and 493 (47.45%) were rejected by physicians (see [Multimedia Appendix 1](#), [Figure S1](#)). The urology, dermatology, chest medicine, family medicine, metabolism, and otolaryngology departments observed higher alert rates of 10.61% (685/6453), 9.80% (633/6453), 6.91% (446/6453), 6.61% (427/6453), 6.52% (421/6453), and 6.50% (420/6453), respectively. Moreover, eight departments, including rehabilitation medicine, infectious disease, and ophthalmology, had alert rates of more than 1%. Gender, diagnosis codes, disease codes, physician IDs, and department IDs were used to develop and validate our prediction model (see [Table 1](#)).

### Prediction Performance of Machine Learning Algorithms

We developed five types of machine learning models to predict physician response. To determine the overall performance of predictive models, six evaluation metrics were applied. Among all the machine learning models, ANN showed the best performance (AUROC 0.94) (see [Figure 4](#) and [Multimedia Appendix 1](#), [Figure S2](#)).

The accuracy of the ANN, RF, NB, GB, and SVM models were 0.88, 0.85, 0.83, 0.82, and 0.57. The sensitivity and specificity of the ANN, RF, NB, GB, and SVM models were 0.87, 0.88, 0.87, 0.79, and 0.57 and 0.83, 0.82, 0.78, 0.90, and 1.0, respectively (see [Table 2](#)).

**Figure 4.** Performance of machine learning prediction models. ANN: artificial neural network; GB: gradient boosting; RF: random forest; NB: naïve Bayes; ROC: receiver operating characteristic; SVM: support vector machine.



**Table 2.** Performance of the prediction models.

Algorithm	Accuracy	Sensitivity, %	Specificity, %	PPV <sup>a</sup> , %	NPV <sup>b</sup> , %	F1
Artificial neural network	0.885	87.01	83.46	87.84	82.40	87.42
Random forest	0.857	88.29	82.48	86.62	84.57	87.44
Naïve Bayes	0.835	87.48	78.74	83.22	84.03	85.29
Gradient boosting	0.828	79.46	90.24	94.59	67.15	86.36
Support vector machine	0.575	57.45	100.0	100.0	0.54	72.97

<sup>a</sup>PPV: positive predictive value.

<sup>b</sup>NPV: negative predictive value.

## Discussion

### Principal Findings

CDSSs directly assist physicians in making correct clinical decisions that ultimately reduce prescription errors by generating real-time alerts and lessen probable unwanted consequences. Clinical workflow is often impaired by excessive numbers of alerts; therefore, physicians pay less attention to alerts and even ignore alerts indiscriminately. This study focused on physicians' recent practice patterns and represented the findings of machine learning models to predict physicians' responses to alerts from a disease medication-related CDSS. The key findings are as follows: (1) an ANN model can correctly predict physicians' responses with higher accuracy than other models and (2) we identified potential features that could provide insight into the system design. These findings may contribute to building a sophisticated provider-friendly interface in which a CDSS may offer real-time alerts if the prediction is positive for that individual physician. If the prediction is negative, that means physicians might not accept the alert; therefore, the CDSS will

not generate alerts during the prescribing of prescriptions or will provide soft or passive alerts without interruption. However, all the alerts would be recorded and the report sent to the individual physician by email on a weekly basis to inform them of how important the alerts were in order to reduce unwanted consequences.

### Clinical Implications

CDSSs have already shown their capability to improve patient safety and quality of care by lowering the number preventable medication errors [32-34]; however, an unreasonable override rate raises questions regarding the quality of CDSSs. Patient safety and effective care could be improved by initiating sophisticated criteria for generating alerts in the CDSS that prevent alert fatigue and minimize the override rate [35-37]. Identification of physicians and departments who override alerts more often would help to reduce the override rate and help us understand how physicians would respond to drug-disease alerts, which would result in immense benefits. There are no previous studies that used a machine learning prediction model to identify physicians and departments who override alerts more often. In

this study, machine learning algorithms were used to reduce alert fatigue by identifying physicians and departments who override alerts more often. Our findings are consistent with existing research that showed physicians played a great role in alert override [38]. Bell et al showed that alert override can be minimized by physicians' preferences for alert selection [39]. There are several reasons that can make physicians override alerts. First, current medication-related CDSSs are not designed to take the patient's previous medication history into account. Sometimes patients are already tolerant of the drug and physicians need to override the alert and prescribe the drug [40]. Second, some CDSSs required an entry for the reason for alert override and that lead to an unacceptable time burden for physicians [41]. Third, physicians believe that *they already know* the alert is inappropriate based on their experience, so they are more likely to override the alert [13,42]. Our study also provides a very important point: no matter how accurate the CDSS is according to the most relevant knowledge base, the alert acceptance was highly affected by the individual physician's perspective. Our model will reduce the gap between real-world clinical practice and knowledge-based theory.

Yeh et al [43] demonstrated that dermatology, gynecology-obstetrics, family medicine, and ophthalmology departments had higher acceptance rates; however, pediatrics, psychiatry, and internal medicine departments, such as cardiology, endocrinology and metabolism, gastroenterology, hematology, rheumatology, and general medicine, had lower

acceptance rates. In our study, we also found that physicians' decisions vary from department to department.

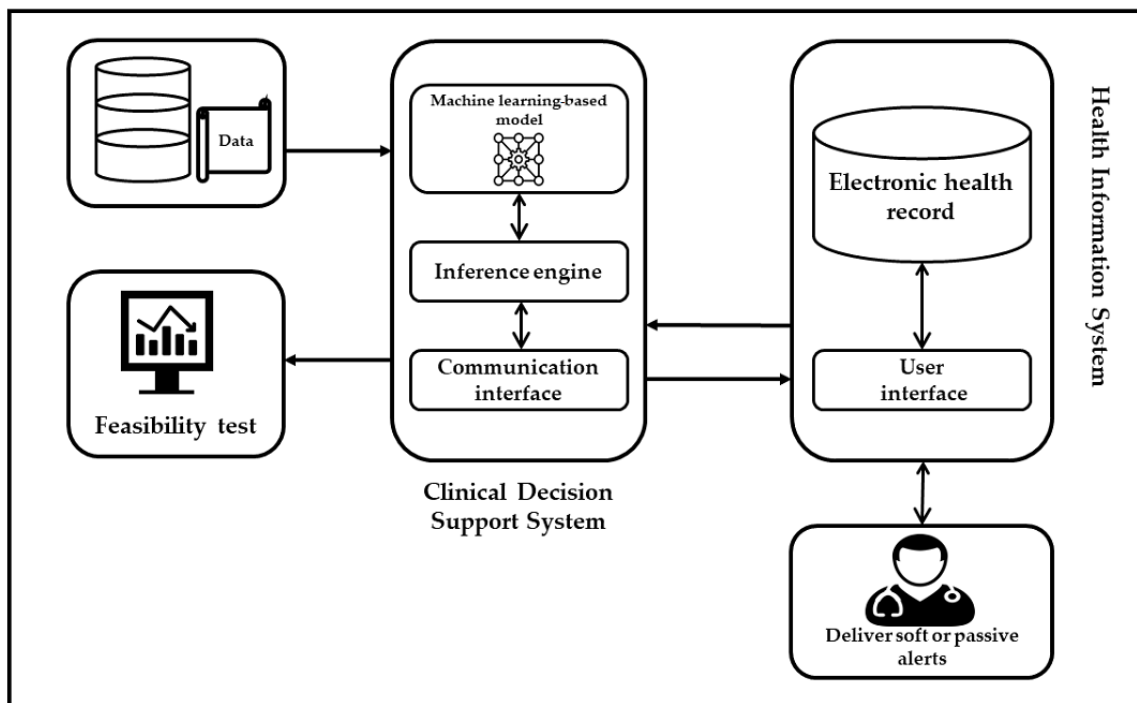
**Strengths and Limitations**

This study has several strengths. First, this is the first study to use machine learning algorithms to predict physicians' intentions to accept or reject alerts. This model may help to reduce alert fatigue in the current CDSS. Second, this study is personalized for each physician. Third, the performance of the model is satisfactory, such that it would help to reduce alert fatigue. Despite several strengths, our study also has several limitations that need to be addressed. First, we did not include free-text override reasons in our analysis, and free-text reasons could add additional value to our model. However, our model provided the AUROC with decent specificity and sensitivity. Second, we did not include physicians' experiences, working periods, ages, and genders in this prediction model. These data are difficult to collect retrospectively because EHR systems do not record this type of information. Third, we have only used one hospital dataset; multiple hospital datasets would make our model more reliable.

**Future Works**

This was the first part of our work. In the future, we will integrate our prediction model into the CDSS in order to check the feasibility of our model. It will help to reduce alert fatigue and result in a sophisticated CDSS by providing soft or passive alerts. Moreover, we will also try to get feedback from physicians about our prediction model (see Figure 5).

Figure 5. Future direction of this study.



**Conclusions**

The findings of the study showed the potential for machine learning prediction models to predict physicians' responses with high sensitivity and specificity. Among the five machine

learning algorithms, the ANN model showed greater performance than the other models. This model can be a promising tool to reduce alert fatigue from CDSSs in clinical settings and can help to correctly identify an individual's alert acceptance rate.



## Acknowledgments

We would like to thank AESOP (AI-Enhanced Safety of Prescription) Technology for giving us data and technological support to conduct this study. This research was funded, in part, by the Ministry of Education (MOE) (grant numbers MOE 109-6604-001-400 and DP2-109-21121-01-A-01) and the Ministry of Science and Technology (MOST) (grant number MOST109-2823-8-038-004).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Description of alerts in different departments and problem fitting checks.

[\[DOCX File, 95 KB-Multimedia Appendix 1\]](#)

## References

1. Kruse CS, Ehrbar N. Effects of computerized decision support systems on practitioner performance and patient outcomes: Systematic review. *JMIR Med Inform* 2020 Aug 11;8(8):e17283 [[FREE Full text](#)] [doi: [10.2196/17283](https://doi.org/10.2196/17283)] [Medline: [32780714](https://pubmed.ncbi.nlm.nih.gov/32780714/)]
2. Campbell E, Guappone K, Sittig D, Dykstra R, Ash JS. Computerized provider order entry adoption: Implications for clinical workflow. *J Gen Intern Med* 2009 Jan;24(1):21-26 [[FREE Full text](#)] [doi: [10.1007/s11606-008-0857-9](https://doi.org/10.1007/s11606-008-0857-9)] [Medline: [19020942](https://pubmed.ncbi.nlm.nih.gov/19020942/)]
3. Monteiro L, Maricoto T, Solha I, Ribeiro-Vaz I, Martins C, Monteiro-Soares M. Reducing potentially inappropriate prescriptions for older patients using computerized decision support tools: Systematic review. *J Med Internet Res* 2019 Nov 14;21(11):e15385 [[FREE Full text](#)] [doi: [10.2196/15385](https://doi.org/10.2196/15385)] [Medline: [31724956](https://pubmed.ncbi.nlm.nih.gov/31724956/)]
4. Eiermann B, Rahmner P, Korkmaz S, Landberg C, Lilja B, Shemeikka T. Knowledge bases for clinical decision support in drug prescribing: Development, quality assurance, management, integration, implementation and evaluation of clinical value. In: Jao C, editor. *Decision Support Systems*. London, UK: IntechOpen; 2010.
5. Coleman JJ, van der Sijs H, Haefeli WE, Slight SP, McDowell SE, Seidling HM, et al. On the alert: Future priorities for alerts in clinical decision support for computerized physician order entry identified from a European workshop. *BMC Med Inform Decis Mak* 2013 Oct 01;13:111 [[FREE Full text](#)] [doi: [10.1186/1472-6947-13-111](https://doi.org/10.1186/1472-6947-13-111)] [Medline: [24083548](https://pubmed.ncbi.nlm.nih.gov/24083548/)]
6. Powers E, Shiffman R, Melnick E, Hickner A, Sharifi M. Efficacy and unintended consequences of hard-stop alerts in electronic health record systems: A systematic review. *J Am Med Inform Assoc* 2018 Nov 01;25(11):1556-1566 [[FREE Full text](#)] [doi: [10.1093/jamia/ocy112](https://doi.org/10.1093/jamia/ocy112)] [Medline: [30239810](https://pubmed.ncbi.nlm.nih.gov/30239810/)]
7. Ko Y, Abarca J, Malone DC, Dare DC, Geraets D, Houranieh A, et al. Practitioners' views on computerized drug-drug interaction alerts in the VA system. *J Am Med Inform Assoc* 2007 Jan 01;14(1):56-64. [doi: [10.1197/jamia.m2224](https://doi.org/10.1197/jamia.m2224)]
8. Poly TN, Islam M, Yang H, Li Y. Appropriateness of overridden alerts in computerized physician order entry: Systematic review. *JMIR Med Inform* 2020 Jul 20;8(7):e15653 [[FREE Full text](#)] [doi: [10.2196/15653](https://doi.org/10.2196/15653)] [Medline: [32706721](https://pubmed.ncbi.nlm.nih.gov/32706721/)]
9. Glassman PA, Simon B, Belperio P, Lanto A. Improving recognition of drug interactions: Benefits and barriers to using automated drug alerts. *Med Care* 2002 Dec;40(12):1161-1171. [doi: [10.1097/00005650-200212000-00004](https://doi.org/10.1097/00005650-200212000-00004)] [Medline: [12458299](https://pubmed.ncbi.nlm.nih.gov/12458299/)]
10. Tao L, Zhang C, Zeng L, Zhu S, Li N, Li W, et al. Accuracy and effects of clinical decision support systems integrated with BMJ best practice-aided diagnosis: Interrupted time series study. *JMIR Med Inform* 2020 Jan 20;8(1):e16912 [[FREE Full text](#)] [doi: [10.2196/16912](https://doi.org/10.2196/16912)] [Medline: [31958069](https://pubmed.ncbi.nlm.nih.gov/31958069/)]
11. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 2006 Mar 01;13(2):138-147. [doi: [10.1197/jamia.m1809](https://doi.org/10.1197/jamia.m1809)]
12. Isaac T, Weissman JS, Davis RB, Massagli M, Cyruik A, Sands DZ, et al. Overrides of medication alerts in ambulatory care. *Arch Intern Med* 2009 Feb 09;169(3):305-311. [doi: [10.1001/archinternmed.2008.551](https://doi.org/10.1001/archinternmed.2008.551)] [Medline: [19204222](https://pubmed.ncbi.nlm.nih.gov/19204222/)]
13. Nanji KC, Slight SP, Seger DL, Cho I, Fiskio JM, Redden LM, et al. Overrides of medication-related clinical decision support alerts in outpatients. *J Am Med Inform Assoc* 2014;21(3):487-491 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2013-001813](https://doi.org/10.1136/amiajnl-2013-001813)] [Medline: [24166725](https://pubmed.ncbi.nlm.nih.gov/24166725/)]
14. Topaz M, Seger DL, Slight SP, Goss F, Lai K, Wickner PG, et al. Rising drug allergy alert overrides in electronic health records: An observational retrospective study of a decade of experience. *J Am Med Inform Assoc* 2016 May;23(3):601-608. [doi: [10.1093/jamia/ocv143](https://doi.org/10.1093/jamia/ocv143)] [Medline: [26578227](https://pubmed.ncbi.nlm.nih.gov/26578227/)]
15. Getty DJ, Swets JA, Pickett RM, Gonthier D. System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *J Exp Psychol Appl* 1995;1(1):19-33. [doi: [10.1037/1076-898x.1.1.19](https://doi.org/10.1037/1076-898x.1.1.19)]
16. Seidling HM, Schmitt SPW, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, et al. Patient-specific electronic decision support reduces prescription of excessive doses. *Qual Saf Health Care* 2010 Oct;19(5):e15. [doi: [10.1136/qshc.2009.033175](https://doi.org/10.1136/qshc.2009.033175)] [Medline: [20427312](https://pubmed.ncbi.nlm.nih.gov/20427312/)]

17. Del Beccaro MA, Villanueva R, Knudson KM, Harvey EM, Langle JM, Paul W. Decision support alerts for medication ordering in a computerized provider order entry (CPOE) system. *Appl Clin Inform* 2017 Dec 16;01(03):346-362. [doi: [10.4338/aci-2009-11-ra-0014](https://doi.org/10.4338/aci-2009-11-ra-0014)]
18. Paterno MD, Maviglia SM, Gorman PN, Seger DL, Yoshida E, Seger AC, et al. Tiering drug-drug interaction alerts by severity increases compliance rates. *J Am Med Inform Assoc* 2009 Jan 01;16(1):40-46. [doi: [10.1197/jamia.m2808](https://doi.org/10.1197/jamia.m2808)]
19. van der Sijs H, Aarts J, van Gelder T, Berg M, Vulto A. Turning off frequently overridden drug alerts: Limited opportunities for doing it safely. *J Am Med Inform Assoc* 2008;15(4):439-448 [FREE Full text] [doi: [10.1197/jamia.M2311](https://doi.org/10.1197/jamia.M2311)] [Medline: [18436915](https://pubmed.ncbi.nlm.nih.gov/18436915/)]
20. Gardner RM, Evans RS. Using computer technology to detect, measure, and prevent adverse drug events. *J Am Med Inform Assoc* 2004 Nov 01;11(6):535-536. [doi: [10.1197/jamia.m1651](https://doi.org/10.1197/jamia.m1651)]
21. Wu C, Hsu W, Islam MM, Poly TN, Yang H, Nguyen P, et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. *Comput Methods Programs Biomed* 2019 May;173:109-117. [doi: [10.1016/j.cmpb.2019.01.013](https://doi.org/10.1016/j.cmpb.2019.01.013)] [Medline: [31046985](https://pubmed.ncbi.nlm.nih.gov/31046985/)]
22. Islam MM, Poly TN, Walther BA, Yang HC, Li Y. Artificial intelligence in ophthalmology: a meta-analysis of deep learning models for retinal vessels segmentation. *J Clin Med* 2020 Apr 03;9(4):1018 [FREE Full text] [doi: [10.3390/jcm9041018](https://doi.org/10.3390/jcm9041018)] [Medline: [32260311](https://pubmed.ncbi.nlm.nih.gov/32260311/)]
23. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of predicting health care utilization via web search behavior: A data-driven analysis. *J Med Internet Res* 2016 Sep 21;18(9):e251 [FREE Full text] [doi: [10.2196/jmir.6240](https://doi.org/10.2196/jmir.6240)] [Medline: [27655225](https://pubmed.ncbi.nlm.nih.gov/27655225/)]
24. Heiat A. Comparison of artificial neural network and regression models for estimating software development effort. *Inf Softw Technol* 2002 Dec;44(15):911-922. [doi: [10.1016/s0950-5849\(02\)00128-3](https://doi.org/10.1016/s0950-5849(02)00128-3)]
25. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002 Dec. URL: <https://cogns.northwestern.edu/cbmgl/LiawAndWiener2002.pdf> [accessed 2020-10-11]
26. Zhang HJA. The optimality of naive Bayes. In: *Proceedings of the Nineteenth National Conference on Artificial Intelligence*. 2004 Presented at: Nineteenth National Conference on Artificial Intelligence; July 25-29, 2004; San Jose, CA.
27. Maron ME. Automatic indexing: An experimental inquiry. *J ACM* 1961 Jul;8(3):404-417. [doi: [10.1145/321075.321084](https://doi.org/10.1145/321075.321084)]
28. Chen Z, Zhang T, Zhang R, Zhu Z, Yang J, Chen P, et al. Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. *Atmos Environ* 2019 Apr;202:180-189. [doi: [10.1016/j.atmosenv.2019.01.027](https://doi.org/10.1016/j.atmosenv.2019.01.027)]
29. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21 [FREE Full text] [doi: [10.3389/fnbot.2013.00021](https://doi.org/10.3389/fnbot.2013.00021)] [Medline: [24409142](https://pubmed.ncbi.nlm.nih.gov/24409142/)]
30. Chen T, He T. xgboost: eXtreme Gradient Boosting. *The Comprehensive R Archive Network*. 2020 Sep 02. URL: <https://cran.r-project.org/web/packages/xgboost/vignettes/xgboost.pdf> [accessed 2020-10-11]
31. Wong HB, Lim GH. Measures of diagnostic accuracy: Sensitivity, specificity, PPV and NPV. *Proc Singapore Healthc* 2011 Dec;20(4):316-318. [doi: [10.1177/201010581102000411](https://doi.org/10.1177/201010581102000411)]
32. Abramson EL, Pfoh ER, Barrón Y, Quaresimo J, Kaushal R. The effects of electronic prescribing by community-based providers on ambulatory medication safety. *Jt Comm J Qual Patient Saf* 2013 Dec;39(12):545-552. [doi: [10.1016/s1553-7250\(13\)39070-9](https://doi.org/10.1016/s1553-7250(13)39070-9)]
33. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: What's the goal? *Acad Med* 2002 Oct;77(10):981-992. [doi: [10.1097/00001888-200210000-00009](https://doi.org/10.1097/00001888-200210000-00009)] [Medline: [12377672](https://pubmed.ncbi.nlm.nih.gov/12377672/)]
34. Légat L, Van Laere S, Nyssen M, Steurbaut S, Dupont AG, Cornu P. Clinical decision support systems for drug allergy checking: Systematic review. *J Med Internet Res* 2018 Sep 07;20(9):e258 [FREE Full text] [doi: [10.2196/jmir.8206](https://doi.org/10.2196/jmir.8206)] [Medline: [30194058](https://pubmed.ncbi.nlm.nih.gov/30194058/)]
35. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, et al. Medication-related clinical decision support in computerized provider order entry systems: A review. *J Am Med Inform Assoc* 2007 Jan 01;14(1):29-40. [doi: [10.1197/jamia.m2170](https://doi.org/10.1197/jamia.m2170)]
36. Blecker S, Pandya R, Stork S, Mann D, Kuperman G, Shelley D, et al. Interruptive versus noninterruptive clinical decision support: Usability study. *JMIR Hum Factors* 2019 Apr 17;6(2):e12469 [FREE Full text] [doi: [10.2196/12469](https://doi.org/10.2196/12469)] [Medline: [30994460](https://pubmed.ncbi.nlm.nih.gov/30994460/)]
37. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of decision support in computerized provider order entry: Systematic literature review. *JMIR Med Inform* 2018 Jan 24;6(1):e3 [FREE Full text] [doi: [10.2196/medinform.7170](https://doi.org/10.2196/medinform.7170)] [Medline: [29367187](https://pubmed.ncbi.nlm.nih.gov/29367187/)]
38. Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. *Arch Intern Med* 2003 Nov 24;163(21):2625-2631. [doi: [10.1001/archinte.163.21.2625](https://doi.org/10.1001/archinte.163.21.2625)] [Medline: [14638563](https://pubmed.ncbi.nlm.nih.gov/14638563/)]
39. Bell H, Garfield S, Khosla S, Patel C, Franklin BD. Mixed methods study of medication-related decision support alerts experienced during electronic prescribing for inpatients at an English hospital. *Eur J Hosp Pharm* 2019 Nov;26(6):318-322 [FREE Full text] [doi: [10.1136/ejhpharm-2017-001483](https://doi.org/10.1136/ejhpharm-2017-001483)] [Medline: [31798854](https://pubmed.ncbi.nlm.nih.gov/31798854/)]
40. Heringa M, Siderius H, Floor-Schreuderling A, de Smet PAGM, Bouvy ML. Lower alert rates by clustering of related drug interaction alerts. *J Am Med Inform Assoc* 2017 Jan;24(1):54-59. [doi: [10.1093/jamia/ocw049](https://doi.org/10.1093/jamia/ocw049)] [Medline: [27107437](https://pubmed.ncbi.nlm.nih.gov/27107437/)]

41. Baysari MT, Tariq A, Day RO, Westbrook JI. Alert override as a habitual behavior: A new perspective on a persistent problem. *J Am Med Inform Assoc* 2017 Mar 01;24(2):409-412. [doi: [10.1093/jamia/ocw072](https://doi.org/10.1093/jamia/ocw072)] [Medline: [27274015](https://pubmed.ncbi.nlm.nih.gov/27274015/)]
42. Nanji K, Seger D, Slight S, Amato M, Beeler P, Her Q, et al. Medication-related clinical decision support alert overrides in inpatients. *J Am Med Inform Assoc* 2018 May 01;25(5):476-481. [doi: [10.1093/jamia/ocx115](https://doi.org/10.1093/jamia/ocx115)] [Medline: [29092059](https://pubmed.ncbi.nlm.nih.gov/29092059/)]
43. Yeh M, Chang Y, Wang P, Li Y, Hsu C. Physicians' responses to computerized drug-drug interaction alerts for outpatients. *Comput Methods Programs Biomed* 2013 Jul;111(1):17-25. [doi: [10.1016/j.cmpb.2013.02.006](https://doi.org/10.1016/j.cmpb.2013.02.006)] [Medline: [23608682](https://pubmed.ncbi.nlm.nih.gov/23608682/)]

## Abbreviations

**AESOP:** AI-Enhanced Safety of Prescription

**ANN:** artificial neural network

**ATC:** Anatomical Therapeutic Chemical

**AUROC:** area under the receiver operating characteristic curve

**CDSS:** clinical decision support system

**CPOE:** computerized physician order entry

**EHR:** electronic health record

**GB:** gradient boosting

**ICD-10-CM:** International Classification of Diseases, 10th Revision, Clinical Modification

**MOE:** Ministry of Education

**MOST:** Ministry of Science and Technology

**NB:** naïve Bayes

**NPV:** negative predictive value

**PPV:** positive predictive value

**ReLU:** rectified linear unit

**RF:** random forest

**SVM:** support vector machine

*Edited by G Eysenbach, R Kukafka; submitted 20.04.20; peer-reviewed by L Zhang, T Goodwin, H Demir, S Sarbadhikari; comments to author 10.09.20; revised version received 12.09.20; accepted 19.09.20; published 19.11.20*

*Please cite as:*

*Poly TN, Islam M, Muhtar MS, Yang HC, Nguyen PA, Li YC*

*Machine Learning Approach to Reduce Alert Fatigue Using a Disease Medication-Related Clinical Decision Support System: Model Development and Validation*

*JMIR Med Inform* 2020;8(11):e19489

URL: <https://medinform.jmir.org/2020/11/e19489>

doi: [10.2196/19489](https://doi.org/10.2196/19489)

PMID: [33211018](https://pubmed.ncbi.nlm.nih.gov/33211018/)

©Tahmina Nasrin Poly, Md.Mohaimenul Islam, Muhammad Solihuddin Muhtar, Hsuan-Chia Yang, Phung Anh (Alex) Nguyen, Yu-Chuan (Jack) Li. Originally published in *JMIR Medical Informatics* (<http://medinform.jmir.org>), 19.11.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.